# Laboratory testbed verification of data post-processing strategies for the WFIRST coronagraph instrument

*Authors:* Neil Zimmerman (1), Laurent Pueyo (2), Rémi Soummer (2), Bertrand Mennesson (3), Rahul Patel (4), Eric Cady (3), Byoung-Joon Seo (3), Ilya Poberezhskiy (3), Marie Ygouf (4)

*(1) NASA-Goddard Space Flight Center*
*(2) Space Telescope Science Institute*
*(3) Jet Propulsion Laboratory, California Institute of Technology*
*(4) Infrared Processing and Analysis Center, California Institute of Technology*

One of the objectives of the WFIRST Coronagraph Instrument is to demonstrate post-processing algorithms for space-based exoplanet imaging data in the regime of very high contrast ratios (< 1E-8). We analyzed two 15-hour time series of laboratory images acquired with prototype coronagraph designs on a vacuum chamber testbed at JPL. These data sets enabled us to test the performance of reference differential imaging (RDI) PSF subtraction as a function of time and Euclidean (L2) speckle pattern distance. Starting from an analytical relationship between speckle pattern stability and classical subtraction gain, we compared the classical subtraction result to those of the KLIP RDI algorithm. In both data sets, KLIP RDI contrast gains are typically 3-4× greater than those of classical subtraction. These results suggest that a diverse reference PSF library will enable a significant gain in sensitivity beyond that suggested by the raw, frame-to-frame speckle pattern stability.

## Summary

- We relate speckle stability to image correlation coefficients and reference differential imaging (RDI) post-processing gains. Stability metrics based on Euclidean ($L^2$) distances between images are simple to define and measure, and they place a useful **lower bound** on the achievable RDI post-processing contrast gain.

- We established a new diagnostic **upper bound** on contrast gain based on direct least-squares inversion. In conjunction with the lower bound on contrast gain determined from Euclidean distance metrics, this offers a convenient formula to bracket the worst and best expectable RDI performance on a given image sequence.

- We applied the KLIP algorithm to measure post-processed RDI contrast gains over long time series of HLC and SPC images acquired on the WFIRST OMC testbed at JPL. In both data sets, KLIP contrast gains typically remain 3-4× greater than those of classical subtraction.


## 1. Image stability, correlation, and post-processing gain

In a crude statistical model, we can represent the raw image as a random vector drawn from a normal distribution. Suppose the variance of this random vector, which sets the detection floor, is due exclusively to the raw speckle intensity pattern, ignoring other noise sources like photon counting noise.

Let $X$ be our time-averaged science image. We have already subtracted its spatial mean, so that $E[X] = 0$. Its variance is $E[X^2] = \sigma^2$ and we assume the underlying, non-stellar astrophysical signal contributes negligibly to this quantity.

$Y$ is our estimate of the star PSF in science image $X$. In classical PSF subtraction, $Y$ is simply the nearest neighbor to $X$, in the $L^2$ or Euclidean sense, among the images in the reference library. In the case of the KLIP subtraction algorithm, the estimate $Y$ is formed from the principal components of a reference PSF library. Either approach can trivially compensate for a scale factor or DC offset between the science target star and the reference star PSFs. Therefore, once we have determined $Y$, subtraction is the only operation remaining to produce the post-processed image where we can search for signals above some threshold, before further analysis with forward-modeled photometry and astrometry.

Intuitively, we expect the stability of the instrument to determine the distance

between *Y* and the star PSF in the science image *X*: with worse speckle stability, the expected $L^2$ distance between $X$ and $Y$ will increase, as will the residual variance in the subtracted image.

We can represent the difference between *X* and *Y* with a zero-mean, independent random vector named *U*, with variance $E[U^2] = \delta^2$:

$$U := X - Y$$

The vector *U* can be interpreted as the change in the speckle pattern due to uncontrolled wavefront drifts. Its RMS value, δ, is the Euclidean distance between *X* and *Y*. It is convenient to normalize this distance to the RMS value of the raw science image, to form a distance metric that is independent of the intensity scale:

$$\frac{\|X - Y\|_2}{\|X\|_2} = \frac{\delta}{\sigma}$$

**The expectation value of the normalized distance $\delta/\sigma$ between two images acquired within some time interval is one possible metric for speckle stability.** The gain in detection sensitivity after subtraction is the inverse of this ratio, $\sigma/\delta$. In other words, the post-processed contrast gain is the ratio of the image RMS values before and after subtraction.

## Key definition

**Throughout this paper we will assess the efficiency of post-processing algorithms using the post-processing gain metric:** $\sigma/\delta$

Another metric that can be used to quantify image stability is the *correlation coefficient* between two images, defined by:

$$\rho_{XY} = \frac{E[XY]}{\sqrt{E[X^2]E[Y^2]}}$$

Using $E[Y^2] = \sigma^2 + \delta^2$, we can express $\rho_{XY}$ in terms of $\sigma$ and $\delta$:

$$\rho_{XY} = \frac{1}{\sqrt{1 + \left(\frac{\delta}{\sigma}\right)^2}}$$

Here we used the fact that $E[XY] = E[X^2]$, since *X* and *U* are independent.

Rearranging for $\sigma/\delta$ allows us to predict the contrast gain after subtracting a reference image with a known correlation coefficient:

$$\frac{\sigma}{\delta} = \frac{1}{\sqrt{1/\rho_{XY}^2 - 1}}$$

The above relationship is exact for the case of classical PSF subtraction, since the residual noise is by definition the component of *X* that is uncorrelated with *Y*. However, when we use principal components to construct the estimate *Y*, the raw frame-to-frame correlations will always underestimate the actual post-processing performance. In the next section, we show how this affects the contrast gain on some example data sets.

# 2. RDI analysis of testbed data

## 2.1. Experiment context and approach

We can use a long sequence of laboratory coronagraph images to test the Reference Differential Imaging (RDI) performance in the presence of a drifting wavefront, and compare the contrast gain against that predicted by raw frame-to-frame statistics. In early 2017, an image sequence of approximately 15 hours duration was recorded for each coronagraph type, Hybrid Lyot Coronagraph (HLC) and Shaped Pupil Coronagraph (SPC).

In each sequence, the deformable mirror (DM) actuator commands were fixed according to an initial satisfactory "dark hole" solution. Over the span of the long data sequence, the speckle pattern is gradually altered and the overall contrast level is degraded. Based on discussions with WFIRST stakeholders, we speculate that the source of this degradation is due to slow transients in the DM actuator response. The origin and detailed behavior of these transients at the hardware level are the subject of a separate investigation by WFIRST project engineers at JPL, and we do not use the datasets to investigate their source. Here we merely use the drift to in the coronagraph response to empirically study the impact of changing high-spatial-frequency wavefront errors on RDI post-processing.

Since the images were acquired in one contiguous sequence on a single lab source, we must arbitrarily split the data between "reference" and "science" subsets. Accordingly, for these RDI trials we assign a subset of frames near the beginning to the reference library. To test RDI performance over a wide range of image correlation levels, we then perform RDI on individual science frames spanning the time series. We performed a similar RDI experiment on SPC lab data in the previous year, however with a much shorter 3-hour time sequence[1].

[1] http://www.stsci.edu/wfirst/technicalreports/WFIRST-STScI-TR1605.pdf

## 2.2. Pre-processed image statistics

**Figure 1** shows the correlation coefficients between all pairs of images in the SPC data set, after they were match-filtered using an off-axis PSF model. The gradual decorrelation of the speckle pattern is evident from the drop in correlation values between images from the beginning and end of the data sequence, reaching down to almost 0.91 in the worst cases.

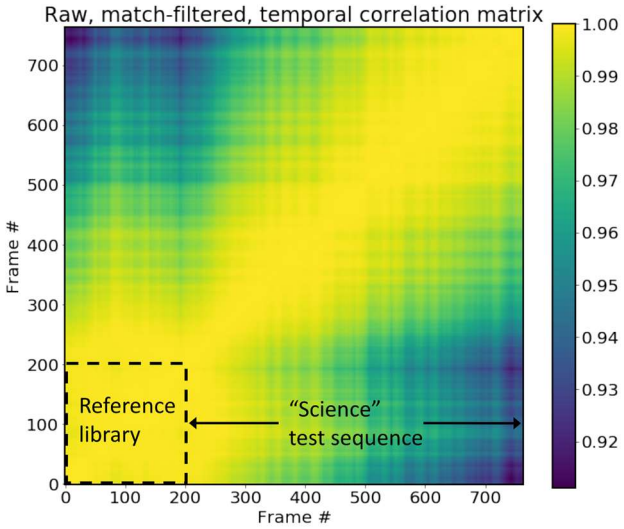



***Figure 1**: Frame-to-frame correlation coefficient matrix of the full SPC image sequence, after applying the PSF matched filter. The first 200 images (4 hours) define the reference library; the remaining frames are treated as individual science frames to test PSF subtraction for various correlation levels.*

**Figure 2** shows the correlation coefficients between each "science" frame and its nearest neighbor (maximum correlation) in the initial 4-hour "reference" library, for both data sets. The rate of decorrelation witnessed in the HLC sequence is noticeably faster.

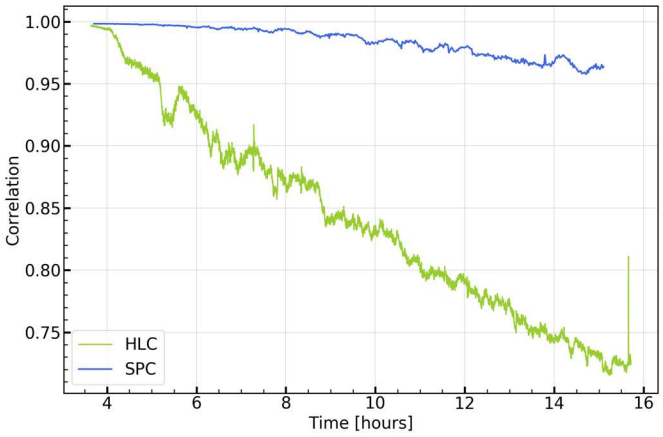



***Figure 2**: Frame-to-frame correlation coefficient between each "science" frame and its nearest neighbor in the reference library. The rate of decorrelation is much higher in the HLC data.*

**Figure 3** displays maps of the temporal RMS variation in each pre-processed image sequence. These maps illustrate the fact that strongly varying speckle features are concentrated in specific regions of the FoV rather than uniformly distributed. Data animations (not shown here) reveal that the changes in the intensity pattern are secular rather than cyclical in nature, consistent with degradation in the wavefront as the deformable mirror actuators drift away from the "dark hole" solution, and similar to trends observed in previous SPC lab data. The goal of the present report is to quantify how much post-processing can mitigate these DM induced starlight suppression drifts.

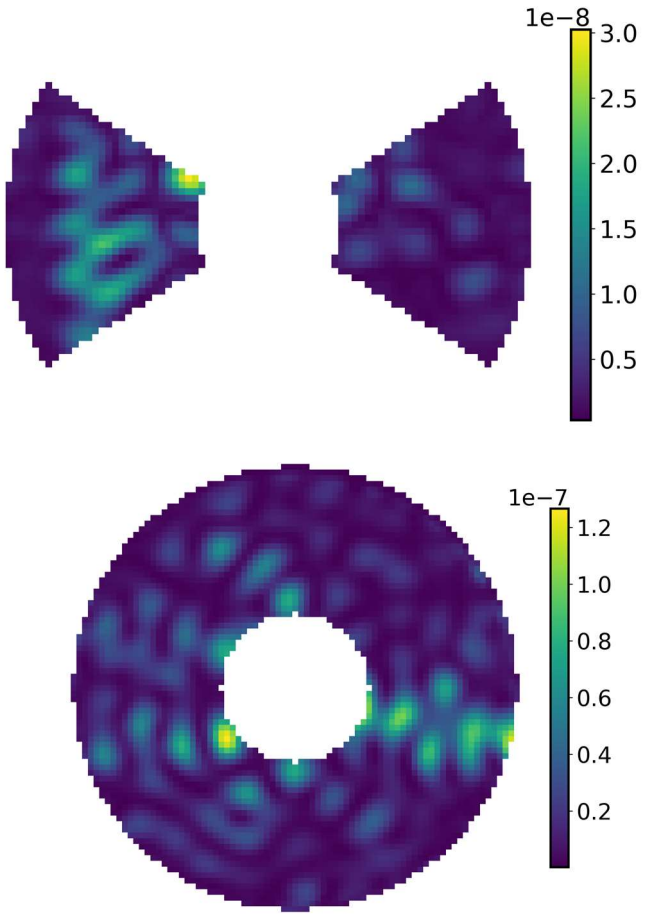



***Figure 3**: Map of the temporal RMS over the full data sequence, measured in each pixel, in units of normalized intensity (contrast equivalent). Top: SPC, Bottom: HLC.*

## 2.3. Post-processing results

**Figure 4** shows the post-processed contrast gains we measured for the SPC and HLC data. We measure the contrast gains as the ratio of spatial standard deviations of the matched-filtered image before/after subtraction, averaged over the field of view. We chose to use matched filter instead of aperture photometry in order to simulate realistic detections scenarios with the SPC coronagraphs (whose elongated PSF with side lobes need to be treated using cross correlation).

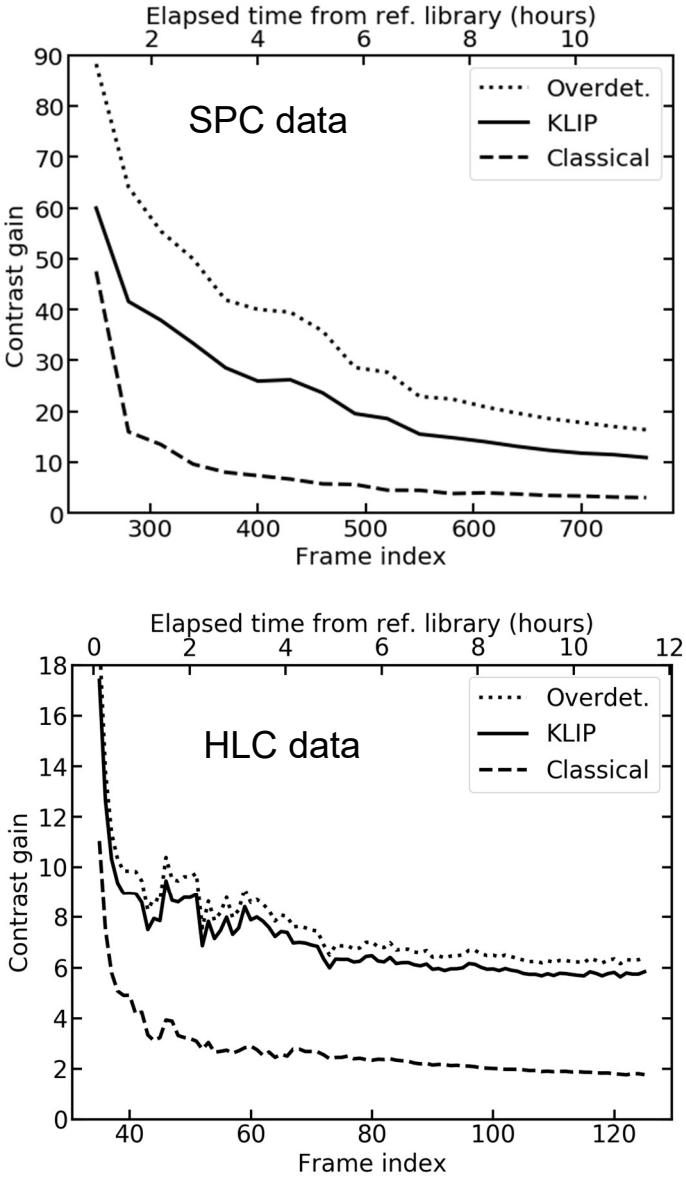



***Figure 4****: Contrast gain for each science frame in the SPC (top) and HLC (bottom) data sets, annotated with elapsed time since the end of the reference library acquisition. The classical gain curve (dashed) indicates the contrast gain after the subtracting the max-correlated reference library image. The KLIP gain curve (solid) indicates the contrast gain after subtracting the PSF estimate determined from the principal components of the 200-frame reference library. The dotted line is the direct, overdetermined least-squares subtraction of reference images, representing an upper bound diagnostic.*

### 2.3.1. Classical subtraction

For the case of classical PSF subtraction, the contrast gains are nearly exactly those predicted by the peak reference library correlation levels for each science frame in Figure 1. For example, in the first science image, frame

#250, acquired approximately 1 hour after the end of the reference library sequence, the peak reference library correlation is 0.9998. Our formula in Section 1 predicts a gain of 50× for this correlation value, which is very near the 47.5× measured in the residual image. At the end of the science sequence, frame #760, when the peak reference library correlation is only 0.946, we expect a contrast gain of 2.9×, again very near the measured gain of 3.0×.

### 2.3.2. KLIP subtraction

Consistent with our previous experience, applying the KLIP algorithm to form the PSF estimate significantly improves the contrast gain. As for the classical case, these contrast gains are determined by the comparing the spatial RMS before and after subtracting the PSF estimate. However, with PCA-based PSF subtraction methods, we must correct for the over-subtraction (or over-fitting) of astrophysical point sources.

Suppose $Z$ is the matrix of eigenimages computed from the reference library, where each column is an eigenimage. And suppose $P$ is the vector PSF model at a given field point, with its spatial mean subtracted. We can then measure the scalar attenuation of a source at this location in terms of matched filter "throughput", based on the inner product between $P$ and $ZZ^TP$, the latter term containing the projection of $P$ onto $Z$:

$$\text{KLIP throughput}, \alpha = 1 - \frac{P^T Z Z^T P}{P^T P},$$

resulting in a corrected post-processing gain expression:

$$E[\alpha]\frac{\sigma}{\delta}$$

where E[$\alpha$] is the mean point source attenuation measured over a large sample of positions in the FoV. Note that this formula only holds perfectly when the reference images do not contain planet signal. This is the case in our RDI scenario, but it is important to note that it is only an approximation in the case of Angular Differential Imaging (ADI). We use an ideal, position-dependent off-axis PSF model specific to each coronagraph to compute E[$\alpha$] (for the bowtie-shaped SPC FoV, 6 angular separations and 10 position angles; for the annular HLC FoV, 9 angular separations and 24 position angles). The resulting contrast gains are plotted in the solid curve in Figure 4. They range from 60× for frame #250 to 11× for frame #760.

### 2.3.3. Direct least-squares subtraction

The third type of curve we plotted in Figure 4 is a diagnostic upper bound on RDI performance. This diagnostic provides a rapid assessment that bypasses the steps in the KLIP algorithm of matrix factorization, basis truncation, off-axis PSF modeling, and over-subtraction calibration. We instead compute the subtracted PSF estimate as a direct least-squares fit from the reference

library. Suppose $R$ is the matrix of reference library images, where each column is a reference image. We represent the fit to science image $X$ as an *over-determined* least-squares problem,

$$R\beta = X.$$

The least-squares coefficient solution vector is given by

$$\hat{\beta} = R^{+}X,$$

where $R^{\dagger}$ is the left pseudo-inverse of the reference matrix, $R^{+} = (R^{T}R)^{-1}R^{T}$. In practice the matrix $(R^{T}R)$ is not guaranteed to be invertible, and the mode cutoff in the KLIP algorithm serves a regularization mechanism. As a consequence using the $\beta$ coefficients based on the inversion of a poorly conditioned matrix will lead to severe signal over-fitting. However, since we are merely estimating an upper bound on the post-processing gain, it is acceptable to use a simple mathematical formula that emulates an algorithm with very low planet throughput. Under this formalism, the PSF estimate is:

$$\begin{aligned}\hat{X} &= R\hat{\beta} \\ &= RR^{+}X.\end{aligned}$$

In our plots in Figure 4, the resulting contrast gain,

$$\frac{\sigma}{\delta} = \frac{\|X\|_2}{\left\|\hat{X} - X\right\|_2}$$

is the dotted curve labeled 'Overdet.' Again we emphasize that this direct least-squares approach runs the risk of severe oversubtraction, which is deliberately *not* accounted for in this purely diagnostic metric. Interestingly, for the HLC data the 'Overdet' contrast gains remain within just a few percent of the KLIP result. This is probably due to lower correlations both within the reference library and between each science image and the reference library. In this scenario the direct least-squares problem is less ill-posed, and thus the un-regularized solution is not plagued by over-subtraction. We checked the ratio of max/min eigenvalues for both the HLC and SPC covariance matrices, and found that it is 2 orders of magnitude lower for the former, which is consistent with our interpretation.

### 2.3.4. KLIP versus classical subtraction as a function of correlation

We plot the contrast gains again in **Figure 5**, this time against reference library correlation coefficient on a scale that is linear in terms of inverse $L^2$ distance metric ($\sigma/\delta$). This serves to demonstrate that the contrast gain measured after classical subtraction matches very closely the value predicted based on the correlation value of the reference frame used to form the PSF estimate (selecting the single reference frame with the maximum correlation).

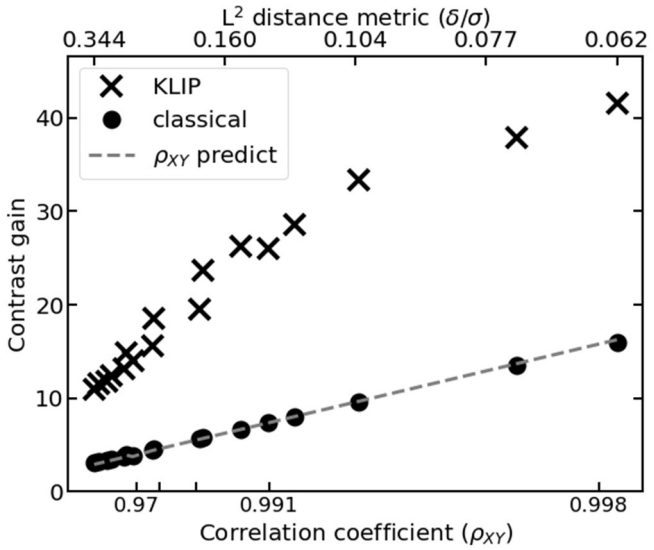



*Figure 5 - Contrast gain in post-processed SPC data as a function of reference library nearest neighbor correlation coefficient (bottom axis), and the equivalent $L^2$ distance metric (top axis). The predicted line (dashed gray) is the value of the inverted correlation formula, solved for $\sigma/\delta$ . The classical contrast gain points (dots) are the measured RMS ratio before/after classical subtraction. The KLIP contrast gain points (crosses) are the measured RMS ratio before/after subtraction, corrected for point source oversubtraction.*

The post-processing advantage of KLIP over classical subtraction is above a factor of 3 for all science frames beginning from frame #340 (2.7 hours elapsed from the final reference image). In **Figure 6**, we plot the ratio of KLIP gain over that of classical subtraction, showing that the advantage factor falls in the range 3 to 4 for all images except for the three most highly correlated frames at the beginning of the trial sequence. The gain ratio remains roughly flat out to the most distant (worst case) images.

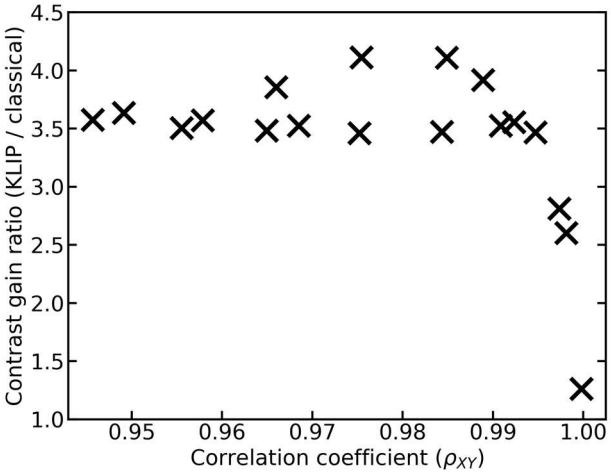



*Figure 6 - Ratio of spatial mean contrast gains achieved with KLIP over classical subtraction, plotted against reference library nearest neighbor correlation.*

## 3. Acknowledgments

This WFIRST study was carried out at Space Telescope Science Institute (STScI) under subcontract No. 1506553, funded by the Jet Propulsion Laboratory (JPL), which is managed for NASA by the California Institute of Technology. Part of this work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA.

The decision to implement the WFIRST mission will not be finalized until NASA's completion of the National Environmental Policy Act (NEPA) process. This document is being made available for information purposes only.

## Appendix A. Matrix algebra KLIP recipe

Here we translate the KLIP algorithm (Soummer, Pueyo, & Larkin, ApJ, 2012; hereafter abbreviated SPL2012) from its textbook formalism to a compact, matrix algebra recipe. The vectorized form of the recipe elucidates two features: (1) the convenient substitute of the reduced SVD factorization for the covariance matrix approach when computing an eigenimage basis; (2) the underlying connection between KLIP and direct, least-squares, LOCI PSF fitting.

First we arrange n spatial-mean-subtracted reference images, each with m pixels, in columns to form the reference image array $R$ (m rows x n columns and m > n). The empirical covariance matrix of the reference images, $R^T R$, is diagonalized by the eigendecomposition:

$$R^T R = V \Lambda V^T$$

Here $V$ is a square matrix consisting of $n$ columns of eigenvectors, and $\Lambda$ is an $n$x$n$ diagonal matrix.

Substituting for the SVD factorization of $R$, $R = U \Sigma V^T$, leads to the relationship between the eigenvalues of the covariance matrix, the singular values of $R$, the eigenvectors of the covariance matrix, and the singular vectors of $R$:

$$\begin{aligned} R^T R &= V \Sigma^T U^T U \Sigma V^T \\ &= V \Sigma^T \Sigma V^T \end{aligned}$$

Therefore, the eigenvectors of $R^T R$ are equal to the right singular vectors of $R$. Furthermore, $\Lambda = \Sigma^T \Sigma$ and each eigenvalue along the diagonal is the square of a corresponding singular value.

The unitary property of $V$, $V^T V = I$, permits us to rearrange the SVD expression for the left singular vectors:

$$RV = U\Sigma$$

and then

$$RV\Sigma^+ = U_n$$

where $\Sigma^+$ is the $n$x$m$ pseudoinverse of the singular value matrix $\Sigma$, and each element value along the diagonal of $\Sigma^+$ is the inverse of its corresponding element in $\Sigma$, and equivalently the inverse of the square root of the corresponding eigenvalue in $\Lambda$. The $n$ subscript on $U_n$ serves to indicate that only the first $n$ columns of $U$ are recovered since the last $m$ - $n$ columns of $\Sigma^+$ are zero columns. In effect, $U_n$ is the left singular matrix for a "reduced" or "thin" SVD factorization in which we restrict ourselves to computing only $n$ singular vectors instead of the larger $m$x$m$ matrix. In the NumPy linear algebra function numpy.linalg.svd(), for example, the reduced SVD option is selected by the specifying the parameter *full_matrices=False*.

One of the critical steps in the KLIP algorithm is the K-L transform of the reference image array, as described in Equation 5 of SPL2012. Summation algebra notation is used there to express the matrix product between the reference data and the covariance eigenvectors ($V$), and the column-wise scaling of each resulting basis vector by the inverse square root of the eigenvalue. Careful comparison to the factorization of the left singular vectors, $RV\Sigma^{+}$, shows that this expression is mathematically identical. In other words, the K-L transformed basis matrix referred to as $Z$ in SPL2012 Eqn 5 can be computed directly as:

$$\begin{aligned} Z &= RV\Sigma^{+} \\ &= U_n \end{aligned}$$

Before truncating the basis, $Z$ has dimensions *m*x*n* (discarding the last *m* - *n* zero columns). Furthermore, **the first *n* columns of *Z* and *U* are equal, so we can obtain *Z* directly from the reduced SVD of *R***.

Now, suppose we have a column vector $y$ representing a science image we wish to fit as an expansion on $Z$, the K-L eigenimage basis. As a regularization strategy to prevent overfitting the non-stellar astrophysical signal in $y$, we truncate $Z$ to $k < n$ basis vectors. Let us call the resulting *m*x*k* basis matrix $Z_k$. By a matrix product, we can project $y$ on $Z_k$ and sum the corresponding linear combination of vectors from $Z_k$ (SPL2012 Eqn 8):

$$\hat{y} = Z_k Z_k^T y$$

The above expression for our PSF estimate, $\hat{y} = Z_k Z_k^T y$ can also be interpreted as a low-rank approximation to the expansion of $y$ as a direct linear combination of reference images. Suppose $\beta$ represents the unknown coefficients that solves the system of linear equations $R\beta = y$. The overdetermined least-squares coefficient solution is $\hat{\beta} = R^{+}y$, where $R$ is the left pseudoinverse ($R^{+} = (R^T R)^{-1} R^T$) and the expansion of $y$ on $R$ is $\hat{y} = RR^{+}y$. If we now expand $R$ and $R^{+}$ by their respective SVD factorizations, we find:

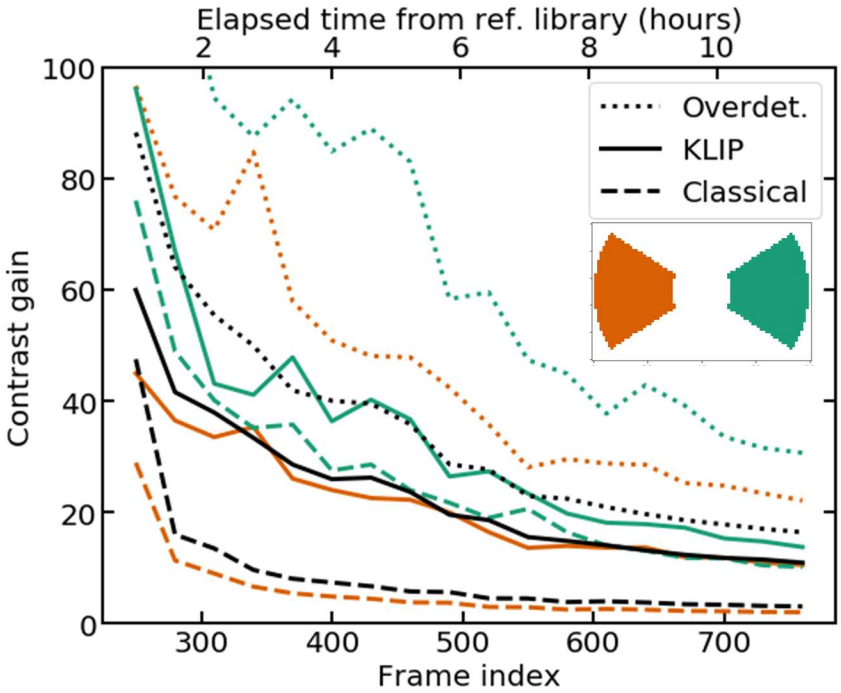
$$\begin{aligned} \hat{y} &= RR^{+}y \\ &= U\Sigma V^T V\Sigma^{+}U^T y \\ &= U\Sigma\Sigma^{+}U^T y \\ &= U_n U_n^T y \end{aligned}$$

In the last line we discarded the last *m* - *n* singular vectors from $U$ due to the fact that the last m - n columns of $\Sigma\Sigma^{+}$ are zero columns. Then **the reference image expansion reduces to the formula for the KLIP PSF estimate without basis truncation**.

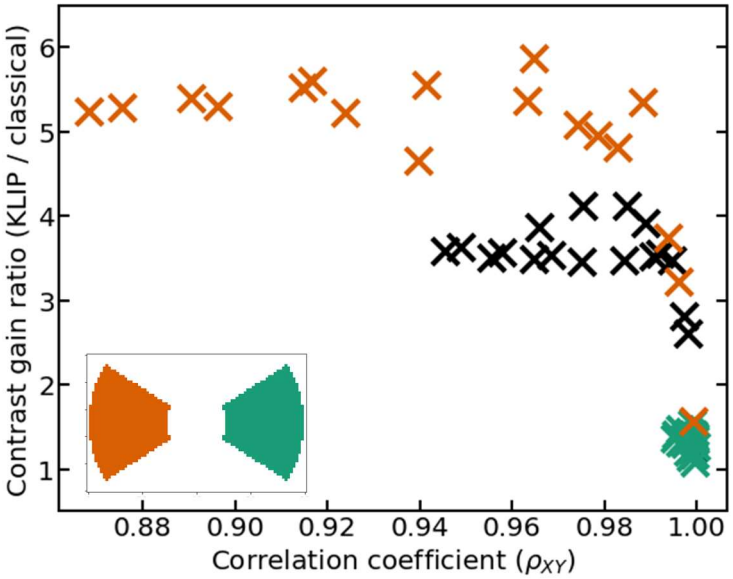
Typically, if we fit our PSF to a linear combination of columns in $R$ and do not apply some form of regularization, as through the KLIP method of basis truncation, then we face a numerically ill-conditioned problem. This introduces the risk of overfitting the astrophysical signal, and furthermore the level of bias will be sensitive to noise in the data. This numerical

instability is the essential disadvantage of LOCI, the least-squares PSF subtraction algorithm first proposed for high-contrast imaging data (Lafreniere et al., 2007).

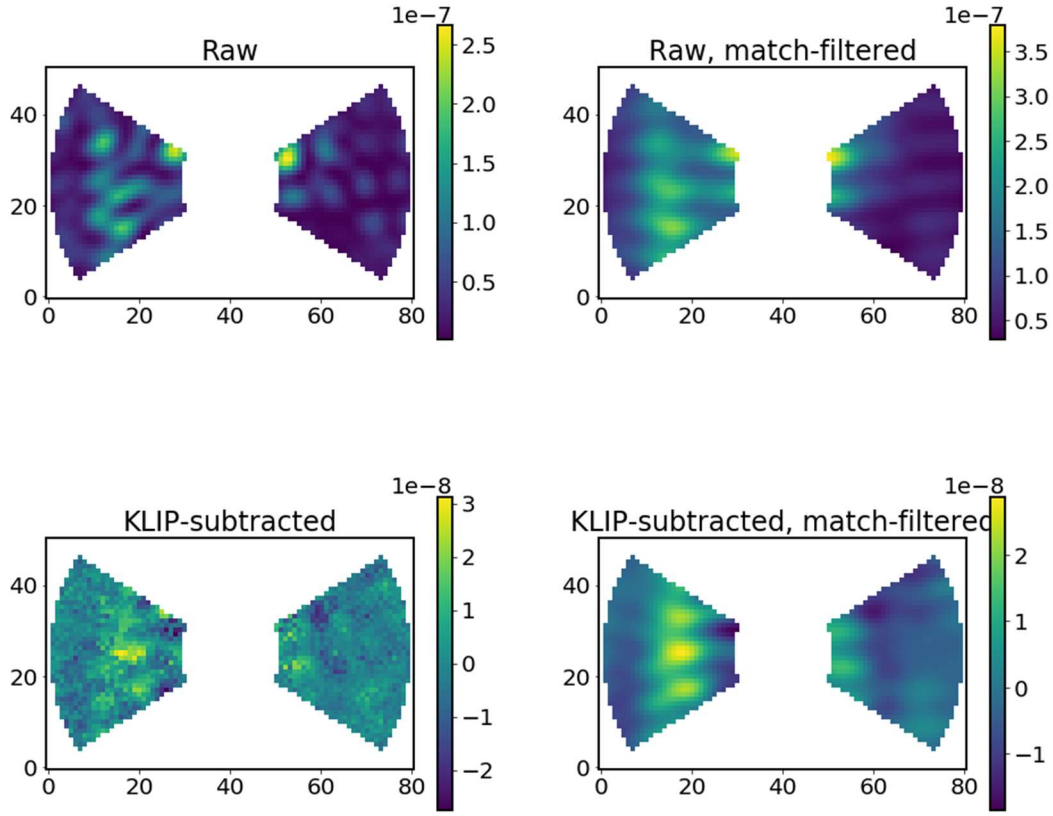
## Appendix B. Supplemental SPC data analysis




*Contrast gain measured independently in each half of the SPC bowtie FoV. The contrast gain is generally higher in the right-hand side of the FoV where the speckle pattern is relatively stable (cf. Figure 3). Another consequence of the high right-hand-side correlations (> 0.995) between science image and reference library is that the classical gain approaches much closer to the KLIP gain.*




*Ratio of spatial mean contrast gains achieved with KLIP over classical subtraction, measured independently in each half of the bowtie FoV, plotted against reference library nearest neighbor correlation.*




*Example KLIP RDI applied to SPC frame number 760 (the last science frame in the trial, with correspondingly lowest reference correlation and worst contrast gain) with a fake planet inserted in the left-hand side of field of view at 3E-8 contrast.*

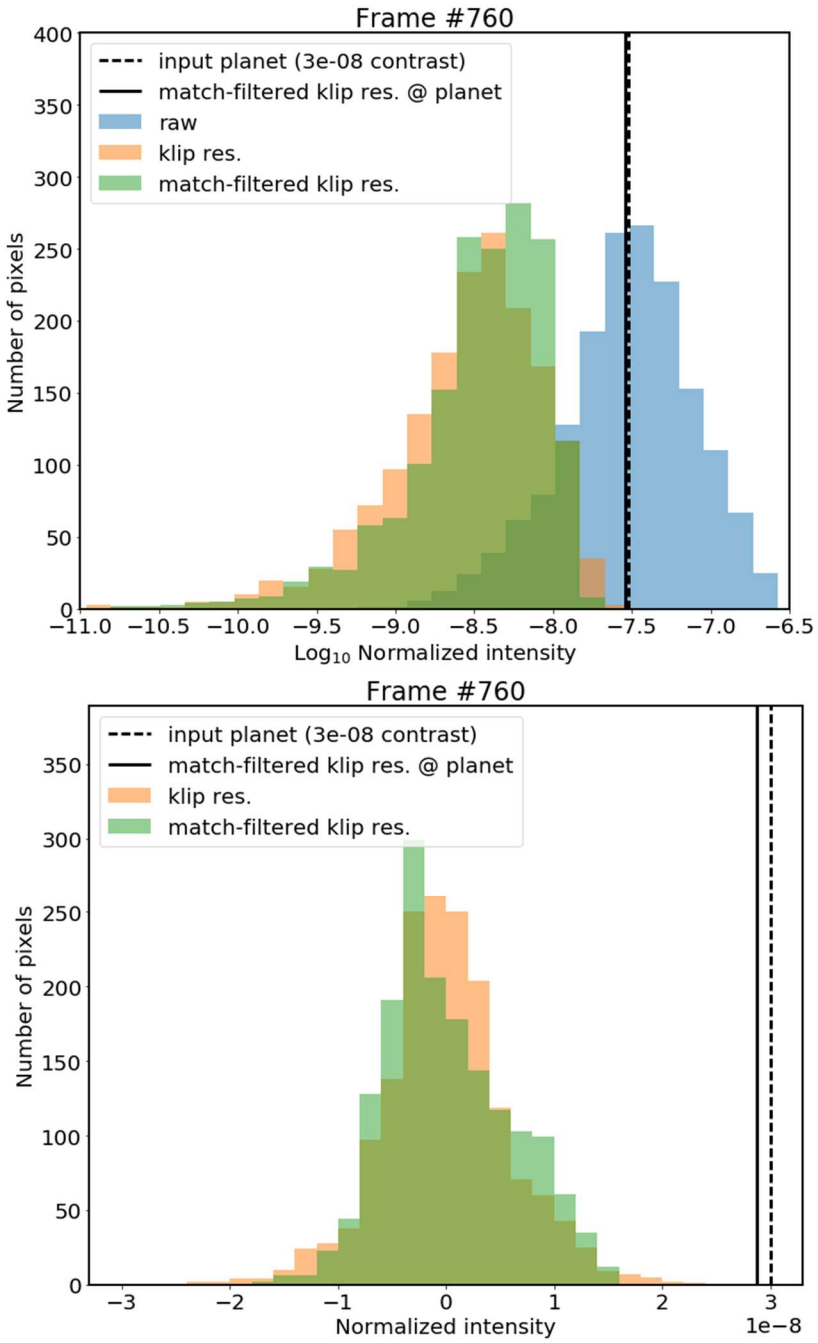



*Figure 8 - Distribution of pixel values in the pre-processed SPC image (blue), after KLIP subtraction (orange), and PSF matched filter (green). The upper histogram shows the distribution on a logairthmic intensity scale, and the bottom histogram shows the residual values on a linear intensity scale. Vertical lines show the flux of the inserted fake planet, and its corresponding peak value in the final KLIP-subtracted, matched-filtered image.*